Questionnaires as interventions: can taking a survey increase teachers' openness to student feedback surveys?

Article *in* Educational Psychology · July 2017

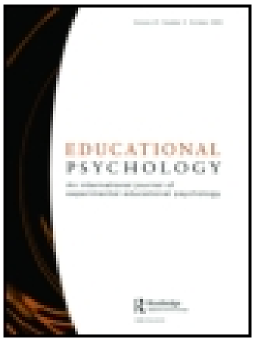
DOI: 10.1080/01443410.2017.1349876

Hunter Gehlbach
University of California, Santa Barbara
46 PUBLICATIONS **1,270** CITATIONS

Carly D. Robinson
Harvard University
16 PUBLICATIONS **120** CITATIONS

Ilana Finefter-Rosenbluh
Monash University (Australia)
11 PUBLICATIONS **65** CITATIONS

Jack Schneider
University of Massachusetts Lowell
32 PUBLICATIONS **216** CITATIONS



Routledge
Taylor & Francis Group

Educational Psychology

An International Journal of Experimental Educational Psychology

ISSN: 0144-3410 (Print) 1469-5820 (Online) Journal homepage: http://www.tandfonline.com/loi/cedp20

Questionnaires as interventions: can taking a survey increase teachers' openness to student feedback surveys?

Hunter Gehlbach , Carly D. Robinson, Ilana Finefter-Rosenbluh, Chris Benshoof & Jack Schneider

To cite this article: Hunter Gehlbach , Carly D. Robinson, Ilana Finefter-Rosenbluh, Chris Benshoof & Jack Schneider (2017): Questionnaires as interventions: can taking a survey increase teachers' openness to student feedback surveys?, Educational Psychology, DOI: 10.1080/01443410.2017.1349876

To link to this article: http://dx.doi.org/10.1080/01443410.2017.1349876

EDUCATIONAL PSYCHOLOGY, 2017
https://doi.org/10.1080/01443410.2017.1349876

Routledge
Taylor & Francis Group

OPEN ACCESS

Questionnaires as interventions: can taking a survey increase teachers' openness to student feedback surveys?

Hunter Gehlbach[a], Carly D. Robinson[b], Ilana Finefter-Rosenbluh[c], Chris Benshoof[d] and Jack Schneider[e]

[a]Gevirtz Graduate School of Education, University of California, Santa Barbara, CA, USA; [b]Harvard Graduate School of Education, Cambridge, MA, USA; [c]Faculty of Education, Monash University, Clayton, Australia; [d]Lathrop High School, Fairbanks, AK, USA; [e]College of the Holy Cross, Worcester, MA, USA

ABSTRACT

Administrators often struggle in getting teachers to trust their school's evaluation practices – a necessity if teachers are to learn from the feedback they receive. We attempted to bolster teachers' support for receiving evaluative feedback from a particularly controversial source: student-perception surveys. For our intervention, we took one of two approaches to asking 309 teachers how they felt about students evaluating their teaching practice. Control participants responded only to core questions regarding their attitudes towards student-perception surveys. Meanwhile, treatment participants were first asked whether *teachers should evaluate administrators* in performance reviews and were then asked the core items about student-perception surveys. Congruent with cognitive dissonance theory, this juxtaposition of questions bolstered treatment teachers' support for using student surveys in teacher evaluations relative to the control group. We discuss the implications of these findings with respect to increasing teacher openness to alternative evaluation approaches, and consider whether surveys show promise as a vehicle for delivering interventions.

ARTICLE HISTORY
Received 18 November 2016
Accepted 29 June 2017

KEYWORDS
Brief interventions; cognitive dissonance; student feedback surveys; questionnaire; teacher assessment

Shakespeare's winter of discontent may well apply to the current sentiment surrounding teacher accountability systems in the United States. Frustrated educational researchers lament that (over) emphasising test-score-based approaches to assessing teachers ignores major confounding factors such as poverty and the complexity of teaching (Berliner, 2013; Good, 2014; Koedinger, Booth, & Klahr, 2013). Teachers worry that they teach a narrower subset of curricula than ever before and that they often must spend, 'substantial instructional time on exercises that look just like the test-items' (Darling-Hammond, 2010, p. 71). To the chagrin of many policy-makers, almost all teachers continue to receive 'proficient' ratings despite principals reporting that the range of teacher competencies is more variable

CONTACT Hunter Gehlbach hgehlbach@education.ucsb.edu

Supplemental data for this article can be accessed at https://doi.org/10.1080/01443410.2017.1349876

(Kraft & Gilmour, 2016). While the discontent is unlikely to turn into glorious summer any time soon, new developments for districts aspiring to fairly evaluate their teachers offer some hope.

Recent research suggests new approaches to assessing teacher quality – in particular, students reporting their perceptions of their teachers – may be a promising component of a teacher evaluation programme (Kane, McCaffrey, Miller, & Staiger, 2013). However, this potential addition to a teacher evaluation system faces a major problem: teacher resistance. Many teachers and their unions oppose integrating student feedback into teacher evaluations (Cromidas, 2012; Decker, 2012). This opposition is understandable – it is far from intuitive that good data might be gleaned from the reports of capricious second graders or surly sophomores. Furthermore, while some forms of evaluation can improve teaching (Taylor & Tyler, 2012), it remains to be seen whether teachers might learn from this kind of feedback. Still, one thing is clear: If teachers consider student-perception surveys to be unfair or biased, the likelihood that their teaching will improve from this feedback seems vanishingly small.

This study tests the effects of a brief intervention designed to nudge teachers' attitudes to be more favourable towards the use of student-perception surveys in evaluating teaching performance.

Broader context of the study

A brief sampling of the scholarship on evaluating teacher effectiveness contextualises the fraught nature of student-perception surveys. In the United States, the adoption of the No Child Left Behind act generated dissatisfaction as teachers garnered nearly universal 'satisfactory' ratings. In response, districts experimented with new evaluation systems. In particular, many districts began to assess their teachers based in part on students' standardised test scores (Steinberg & Donaldson, 2016). Research suggested that more effective teachers in early grades (as measured through this test-score approach) impacted a host of long-term student outcomes such as pregnancies and lifetime earnings (Chetty, Friedman, & Rockoff, 2011). Consequently, enthusiasm for these evaluation methods grew. Simultaneously, scepticism and critique of this approach erupted (Baker et al., 2010). Some argued that because of the complexity of teaching (Koedinger et al., 2013), students' standardised test scores should only comprise a part of teachers' evaluations – leaving open the question of what other data might provide useful feedback on teachers' effectiveness.

Based in large part on the findings from the *Measures of Effective Teaching* study, student-perception surveys gained traction as a potentially useful component of teacher evaluation systems. The study's authors found that students' perceptions were not only reliable, but possibly more accurate in predicting gains in student learning than observation protocols (Kane et al., 2013). Others found that student surveys about their teachers better predicted student scores on criterion-referenced tests than teacher self-ratings and principal ratings (Wilkerson, Manatt, Rogers, & Maughan, 2000). Additionally, student surveys remain relatively cheap and easy to administer. Perhaps most importantly, these surveys can potentially capture a much richer array of desired teacher qualities than might gleaned from students' test scores (Ferguson, 2012).

However, this idea was hardly less controversial than evaluating teachers on their students' test scores. Which aspects of teaching might students reasonably report on? At which grade levels? For all courses or just academic ones? Should stakes be attached to these surveys – possibly causing students to misreport their true feelings – or should the surveys solely be used to drive improvements in teaching?

Thus, those interested in improving teacher evaluations faced a tough choice. On the one hand, preliminary studies suggested that student reports might be an important, straightforward way to expand our approaches to evaluating teachers (Kane et al., 2013; Wilkerson et al., 2000). On the other hand, if teachers were not open to this approach, it seemed unlikely that the system would work well or that teachers would learn much from the student feedback. Consequently, researchers and district administrators interested in the viability of student-perception surveys as part of teacher evaluations faced a Catch-22: They needed teachers to be open to use student-perception surveys as a part of their evaluation systems. Only then could researchers fairly adjudicate whether student-perception surveys might work as a component of these evaluation systems.

For the sake of the present research, two key points should be remembered. First, the controversial topic of student-perception surveys has emerged within larger controversies surrounding teacher evaluation. So when asked about their attitudes towards student-perception surveys, teachers likely have thought about the issue and may well have strong opinions, i.e. they are unlikely to be blank slates.

Second, these surveys are already happening across the United States and internationally, so school leaders need to get teachers bought-in to learning from student feedback. Numerous states now encourage the use of student-perception surveys to assess K-12 teachers (The Colorado Education Initiative, 2015; MET Project, 2012; TEAMTN, 2015). As Steinberg and Donaldson (2016) report, 17% of the largest US districts employ student-perception surveys in some way. Thus, students are already generating vast quantities of feedback. The question is whether teachers will learn from it. If an intervention could nudge teachers to be slightly more open to learning from this feedback, the resulting effects could improve teaching across much of the United States. Particularly because teaching is so context-dependent – what works for one group of students may or may not translate to the next class period, the next day, or the following year's class – getting feedback that is specific to a particular group of students is vital for teachers.

Leveraging cognitive dissonance through surveys

Our intervention leveraged the social psychological principle of cognitive dissonance (Festinger, 1962). Cognitive dissonance research has been one of the most robust and influential areas of inquiry within social psychology (Brehm, 2007). Current dissonance scholars largely agree that this psychological state arises when individuals experience tension between inconsistent cognitions. Because people desire internal consistency, experiencing incompatible cognitions causes discomfort. In most situations, this uncomfortable tension motivates action to alleviate the tension (Brehm, 2007; Gawronski, 2012; Martinie, Milland, & Olive, 2013). Numerous experiments show that people employ a range of strategies to mitigate this discomfort: by changing one of their beliefs or attitudes, through recalibrating the importance of the relevant cognitions, by engaging in a new behaviour, through changing their ongoing behaviour, or by feeling less responsible for their behaviour (Martinie et al., 2013).

Much of the work on dissonance has focused on the alignment of cognition and behaviours. For instance, Harmon-Jones, Harmon-Jones, and Levy (2015) describe three main paradigms of cognitive dissonance research, each of which implicate a person's behaviours: *induced compliance*, *decision-making*, and *effort justification* studies. From this perspective on cognitive dissonance, 'the negative affective state of dissonance is aroused not by all cognitive conflict but, specifically when cognitions with *action implications* conflict with each other making it difficult to act' (Harmon-Jones et al., 2015, p. 185).

However, others suggest that behaviours or actions may not be required for individuals to experience dissonance. In this view, inconsistent cognitions may serve as a cue for the presence of errors in one's belief system (Gawronski, 2012). Thus, an intriguing question – and one with important practical implications – becomes whether attitude change might be sparked through inconsistent cognitions even if the thoughts have little potential to influence behaviour.

Social psychologists have applied the basic idea of cognitive dissonance across an array of real-world settings to generate a variety of interventions. Through 'foot-in-the-door' techniques, participants find that it becomes much harder to say no to someone after having already made a small concession or done a modest favour (e.g. Freedman & Fraser, 1966). In 'saying-is-believing' interventions, participants publicly espouse a point of view and then subsequently tend to endorse that point of view more strongly (e.g. Aronson, Fried, & Good, 2002; Walton & Cohen, 2011). In other words, to say one thing and believe another would be inconsistent. In these field-experiments the dissonant cognitions again tend to implicate actions.

Before dissonance theory came to the fore in social psychology, scholars in other fields utilised people's desire for internal consistency to demonstrate biased responding in questionnaires. For example, in the late 1940s asking Americans whether communist reporters should be allowed to report on visits to the United States garnered little endorsement (37% of respondents say 'yes'). However, first asking whether US reporters should be allowed to report on the Soviet Union (an idea most everyone endorsed) and *then* asking about the communist reporters dramatically shifted endorsements to 73% (Dillman, Smyth, & Christian, 2014). In this instance, presumably the respondents felt awkward about maintaining a double-standard for Soviet and US reporters and thus shifted their opinions. Thus, experimental evidence exists that is congruent with a cognitive dissonance explanation, even though no actions are implicated. However, one could argue that most respondents have no personal stake in what happens to reporters of different nationalities. Therefore, they might be motivated only by presenting themselves consistently to the administrator of the survey. Because the content of the cognitions is not particularly relevant at a personal level, participants are unlikely to have held strong opinions about these reporters previously. Consequently, changing one's opinion on this issue seems relatively cost-free. To the extent that dissonance occurs at all, it is likely a weak version that might be easily resolved.

The situation becomes more intriguing when we shift to a case that has personal relevance (but no action implications) for survey respondents. In this instance, we might anticipate more strongly held prior attitudes that would be correspondingly harder to shift. In other words, can cognitive dissonance still be sparked by attitudes alone when respondents are personally invested in an issue? This is exactly the case we examined.

The present study

We applied this same psychological principle of cognitive dissonance to the challenge of cultivating teachers' support for using student-perception surveys as a component of teacher evaluations. We randomly assigned a group of teachers to respond to survey questions about their support for student-perception surveys under one of two contexts. Control teachers simply took a five-item survey scale assessing their feelings towards student-perception surveys as the initial part of their survey. Treatment teachers answered the same items, but did so *after* first responding to a parallel scale about teachers evaluating their administrators.

As described in our *Statement of Transparency,* we anticipated that most teachers would endorse their own capacity to capably evaluate their administrators. These relatively high ratings would then spark a sense of dissonance when teachers next answered the items regarding students evaluating teachers. In other words, we anticipated that teachers in the treatment group would think something akin to: (1) Yes, teachers are capable of evaluating and giving feedback to their administrators, (2) I am a fair person, who does not hold double-standards; I am not a hypocrite, and (3) Although some students might be too young, if it is reasonable for teachers to evaluate administrators, it should be reasonable for students to evaluate their teachers.

Congruent with recent best practices for experimental studies (Gehlbach & Robinson, manuscript under review; Simmons, Nelson, & Simonsohn, 2011), we submitted our Statement of Transparency using *Open Science Framework* and pre-registered our main hypothesis that: Treatment teachers will report greater support for student-perception surveys on our five-item composite than their control counterparts (controlling for their status a national- or state-level award winning teacher). Increasingly, scholars have raised concerns about 'researcher degrees of freedom' in which investigators engage in various practices that have problematic repercussions. On the one hand, some of these practices, such as testing numerous covariates, can provide an exhaustive sense of what a data-set might tell us about a particular population. On the other hand, the practices enumerated by Simmons et al. (2011) all serve to inflate the *p*-value of any given analysis. By pre-registering our analysis plan and specifying the model we fit ahead of time, we avoid this concern. By describing a set of exploratory analyses, we also hope to gain additional insights that might be generated from the data-set. Readers should have more faith in the findings corresponding to the pre-registered analysis and should treat the exploratory analysis as hypothesis generating. Finally, we report our findings using confidence intervals and effect sizes rather than relying on null-hypothesis significance testing (Cumming, 2014; Thompson, 1996).

Methods

Participants

We recruited participants through snowball sampling using teachers from a prominent teacher organisation as our initial base of participants. Specifically, we partnered with the National Network of State Teachers of the Year (NNSTOY, www.nnstoy.org), an organisation of teachers who were selected as finalists or winners of State or National Teachers of the Year competitions across the US. In addition to their broad geographic representation, we decided to start from this sample of NNSTOY teachers based on the potential implications of our study. We were especially interested in whether this intervention might work with teachers who were leaders in their respective school communities. If school administrators could use this approach successfully to get buy-in from the leaders in their school, we expected that other teachers might be more likely to be persuaded.

The study focused on K-12 teachers at the end of the 2014–2015 school year. Of the 407 teacher participants who clicked into the survey, 309 participants (n = 157 control; n = 152 treatment) continued the survey long enough to complete the intervention and primary dependent measure (i.e. control participants completed Support for Student-Perception Surveys scale and treatment participants completed both scales). No participants who began the intervention dropped out before completing the primary dependent measure; thus, there was no differential attrition for the treatment participants simply because they had to complete five extra items. Of the 279 participants who completed the entire survey (i.e. all the way through the demographic questions at the end of the survey), 76% were female and 32% were members of the NNSTOY. In terms of race/ethnicity, 85% of participants identified as white or Caucasian, 5% Latino, and less than 5% each for teachers who categorised themselves as African-American, American Indian/Alaskan Native, Middle Eastern, or 'Other.'

Participants taught in 44 states and the District of Columbia, and teachers from all grades, K-12, were represented. Approximately 50% of teachers reported having taught high school in the prior year, 24% taught middle school and 26% taught elementary school. The average amount of teaching experience was 18 years, with a standard deviation of 8.2 years and a maximum of 39 years.

Thus, the sample was relatively representative of the US population of teachers on dimensions such as race and gender – the overall teaching population for 2011–2012 was 82.7% white and 76.2% female (National Center for Educational Statistics, 2013). However, the large proportion of award winning teachers, high numbers of high school teachers and substantial years of experience were not representative of the broader population of teachers. Given the experimental design, the extent to which these discrepancies limit the generalisability of the results is unclear.

Measures

Our measure of Support for Student-Perception Surveys consisted of a five-item scale (α = .86) to assess teachers' views of using student-perception surveys to evaluate teachers. After correlating the errors for items 2 and 3, a confirmatory factor analysis showed that the data fit a one-factor model ($\chi^2_{\text{df} = 309} = 5.89$, $p = .21$; CFI = .997; RMSEA = .039). This measure included questions such as, 'Overall, to what extent is it a good idea to have teachers' performance reviews be partially based on student input?' Both treatment and control participants completed this scale. See Table 1a for item-level descriptive statistics on this measure.

Only treatment participants completed the Support for Teacher-Perception Surveys measure – a five-item scale ($\alpha = .75$) that mirrored the student-perception survey scale and assessed teachers' views of using teacher-perception surveys to evaluate administrators. See Table 1b. After correlating the errors for items 2 and 5, a confirmatory factor analysis showed that the data fit a one-factor model ($\chi^2_{\text{df} = 151} = 5.36$, $p = .25$; CFI = .993; RMSEA = .048). This measure included questions such as, 'Overall, to what extent is it a good idea for administrators' evaluations to be based partially on teacher input?'

Beyond these findings regarding the reliability and structural validity (Messick, 1995) of each scale, acquiring additional indicators of validity was challenging because we developed both scales explicitly for this project. However, we took seriously the notion that validity

Table 1a. Descriptive statistics for support for student-perception survey scale (unadjusted mean, SD, and Pearson (*r*) correlations).

	Treatment		Pearson *r* correlations						Control	
	M	SD	1	2	3	4	5	6	SD	*M*
1) Fair	2.84	1.04	–	.48	.50	.52	.71	.84	1.08	2.47
2) Useful	3.85	.97	.57	–	.54	.25	.47	.73	1.12	3.86
3) Objective	2.68	.97	.70	.60	–	.41	.53	.77	.95	2.63
4) Others	2.12	.98	.61	.41	.56	–	.51	.66	.80	1.83
5) Good idea	2.64	1.16	.78	.56	.70	.61	–	.84	1.05	2.18
6) Overall composite	**2.83**	**.85**	**.89**	**.75**	**.86**	**.77**	**.89**	–	**.77**	**2.60**

Notes: 1) *N*s = 152 for Treatment; 157 for Control.
2) The observed range for each item and the composite were 1 through 5.
3) All correlations are significant at the $p < .05$ level.
4) Intra-scale correlations are below the diagonal for treatment and above the diagonal for control.

Table 1b. Descriptive statistics for the teacher-perception survey scale (unadjusted mean, SD, and Pearson (*r*) correlations).

			Pearson *r* correlations					
	M	SD	1	2	3	4	5	6
1) Fair	3.38	1.09	–					
2) Useful	4.43	.71	.32	–				
3) Objective	3.07	1.07	.55	.19	–			
4) Others	3.30	.98	.59	.14	.51	–		
5) Good idea	3.84	.96	.37	.45	.29	.27	–	
6) Overall composite	**3.61**	**.68**	**.83**	**.53**	**.75**	**.73**	**.66**	–

Notes: 1) *N* = 151–152.
2) The observed range for each item 1 through 5, except for 'useful' (2 through 5); the overall composite was 1.6 through 5.
3) All correlations in the table (except for the Others-by-Useful correlation) are significant at the $p < .05$ level.

should be built into each measure from the outset of the scale development process (Gehlbach & Brinkworth, 2011). As such, we reviewed the literature on both topics, solicited input from numerous teachers about both scales, synthesised these two distinct sources of information, and adhered to standard best practices in survey design in writing the items (Dillman et al., 2014; Gehlbach & Brinkworth, 2011, steps 1–4 of their survey design process). Feedback from a pilot allowed us to revise the scales. We present the final versions of both measures in the Appendix 1.

Frequently, the claim of a scale being 'validated' rests upon a series of correlations with other measures which show particular patterns of convergent and discriminant validity. For example, we take the fact that our two scales correlated moderately ($r_{152} = .52$, $p < .001$) as evidence that they are measuring related concepts as expected (i.e. both are tapping into a general attitude towards feedback surveys). To our knowledge though, no other similar measures of these constructs exist making it challenging to enact this traditional approach to establishing validity. Furthermore, in actuality, validity is not an achieved state but an ongoing process (Gehlbach, 2015). Thus, for newly developed scales we feel as though we have reasonable preliminary evidence of construct validity, though this will be an important area to build upon through future research.

Finally, we also collected demographic data and information on the participants' teaching career at the end of the survey.

Procedures

Through the NNSTOY network, we recruited teachers via emails and posts on social media outlets. We encouraged the NNSTOY participants to take the survey themselves and then to email the survey link to their fellow teachers in their schools and professional networks. Participants were given the opportunity to win a $100 gift card in a lottery. Towards the end of the survey, participants answered open-ended questions and could sign up for future interviews/focus groups to discuss student-perception surveys as part of an ongoing, complementary study.

The survey, administered via Qualtrics, took 5–10 min to complete and remained open for two weeks in June of 2015. After participants completed their consent forms, the Qualtrics platform randomly assigned them to treatment and control. All participants were told that schools and districts across the country are considering using perception surveys as part of performance reviews for teachers, and researchers wanted to get teachers' input on this practice. For control group participants, they then answered the five-item scale regarding their views about the use of student-perception surveys to evaluate teachers.

Before being asked about student-perception surveys, participants in the treatment condition were first told that schools and districts across the country are considering using *teacher* perception surveys as part of performance reviews for administrators, and researchers wanted to get teachers' perspectives on this idea. They then answered the five-item scale regarding their views about the use of teacher-perception surveys to evaluate administrators.

Analytic approach

As noted in our Statement of Transparency, we evaluated our hypothesis using ordinary least-squares regression with NNSTOY status as a covariate:

$$\text{Outcome}_i = \beta_0 + \beta_1 \text{Treatment}_{1i} + \beta_2 X_{2i} + \varepsilon_i,$$

where Treatment_{1i} indicates whether teacher *i* was exposed to the cognitive dissonance treatment or not, X_{2i} is a dummy variable indicating whether the teacher was a member of NNSTOY or not, and ε_i is a residual. We included NNSTOY as a covariate because we assumed that teachers who received such positive, public acclaim for their teaching would be more confident teachers and more open to feedback from students than their non-NNSTOY peers. We hoped the covariate would sharpen the precision of our estimates by accounting for this additional source of variation.

As a first step in our analyses, we checked for violations of random assignment with respect to teachers' gender, race, NNSTOY status, level of schooling taught or years of teaching experience. Second, as a manipulation check, we examined whether teachers generally endorsed the notion that they were competent to evaluate their administrators.

Results

Preliminary analyses

We found no evidence that our random assignment produced non-equivalent groups, Specifically, the treatment and control groups appeared similar with respect to the distribution of: males and females, $\chi^2_1 = 1.03$, $p = .31$; NNSTOY membership, $\chi^2_1 = .07$, $p = .79$; different racial and/or ethnic backgrounds, $\chi^2_5 = 5.76$, $p = .33$; grade-level taught (i.e. elementary, middle, or high school), $\chi^2_2 = 2.00$, $p = .37$; or years of teaching experience, $M_{\text{control}} = 18.43$, SD = 8.43 versus $M_{\text{treatment}} = 17.37$, SD = 8.13, $t_{277} = 1.07$, $p = .29$.

Next, our manipulation was predicated on the assumption that teachers would feel competent to provide objective, fair and useful feedback to their administrators. Had they felt they could not competently provide administrators with feedback, no dissonance would be aroused by concluding that students could not reasonably provide teachers with feedback either. Our assumption appears reasonable. Teachers' mean rating of 3.6 (SD = .68) on the Support for Teacher-Perception Surveys scale is closer to the 'quite' than to 'moderately' response options on the scale. For example, in response to being asked 'Overall, to what extent is it a good idea for administrators' evaluations to be based partly on teacher input?' teachers' mean response was closest to the 'quite a good idea' anchor.

Pre-specified hypothesis test

With these preliminary findings in mind, we tested our primary hypothesis: that our intervention would nudge teachers' opinions about student-perception surveys in a positive direction. As predicted, we found that teachers in the treatment condition supported student-perception surveys more than their control counterparts while controlling for participants' NNSTOY status ($B = .23$, SE = .10, CI: .04, .42). These between-group differences correspond to an effect size of $\beta = .14$, or Cohen's $d = .28$. The confidence interval excludes 0, indicating that the difference between the group means is statistically reliable. Figure 1 shows the means and 95% confidence intervals. As noted by Cumming (2014), overlapping confidence intervals should not be confused as being equivalent to a 'non-significant' result, 'If the two groups' CIs overlap by only a moderate amount … approximately, p is less than

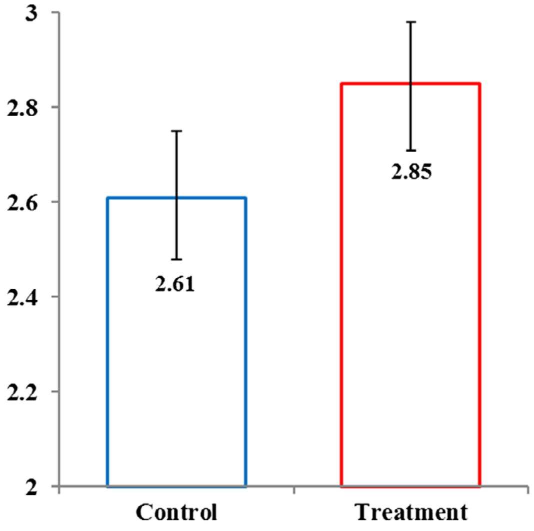


Figure 1. Mean differences and 95% confidence intervals for Support for Student-Perception Surveys by condition controlling for whether teachers were members of the National Network of State Teachers of the Year (or not).

.05' (p. 13). On average then, the treatment teachers were close to 'moderately' endorsing the idea of student-perception surveys while the control teachers were about half-way between the 'mildly' and 'moderately' response options.

Exploratory analyses

We conducted three main types of exploratory analyses – analyses that should be viewed as hypothesis generating or suggestive. The first set of these additional analyses helped us better understand our results and place them into context. Toward this end, we first re-ran our equation testing our core hypothesis without the NNSTOY covariate. Removing teachers' NNSTOY status made essentially no difference ($B = .23$, SE = .09, CI: .05, .41; $\beta = .14$). Similarly, when we included the grade-level taught as a covariate in our original equation, the treatment effect was essentially unchanged ($B = .24$, SE = .10, CI: .05, .43; $\beta = .14$).

We also wanted to know whether teachers' support of student-perception surveys differed based on whether or not they were NNSTOY members. We anticipated that NNSTOY teachers probably received more positive feedback from students (and others) over time and thus might be more open-minded about having their teaching practice evaluated by students. To investigate this possibility, we regressed the Support for Student-Perception Surveys composite on teachers' NNSTOY status. Congruent with our assumption, we found that NNSTOY teachers were more supportive of student-perception surveys than teachers who have not received this recognition ($B = .41$, SE = .10, CI: .21, .62; $\beta = .23$). Similarly, we expected that teachers of earlier grades would be more sceptical that their younger students would have the capacity to provide trustworthy evaluations (as compared to teachers of older students). We explored this assumption by regressing the Support for Student-Perception Surveys composite on the (average) grade-level that teachers taught. Teachers of younger students were, in fact, less likely to endorse student-perception surveys, ($B = .04$, SE = .01, CI: .01, .06; $\beta = .18$).

Finally, Table 1a reveals that the treatment and control groups did not diverge on all items. Specifically, both groups' scores were similar on the utility of student evaluations and the potential for students to be objective; by contrast, bigger differences appeared to emerge for the 'fairness' and 'good idea' items.

The second set of exploratory analyses reflect our attempt to learn more about the plausibility of cognitive dissonance as the hypothesised mechanism driving the group differences. Presumably, for the Support for Teacher-Perception Surveys scale to influence treatment participants on the Support for Student-Perception Survey scale, their responses – at both the item and scale levels – should be correlated. Moreover, one might imagine that the correlation between the parallel items from each scale that invoked implicit comparisons might be higher than the correlation between parallel items that do not invoke such comparisons. For example, the 'fairness' item might invite respondents to think about whether an activity that is fair for teachers to do would also be fair for students.

As shown in Table 2, each parallel item and the overall scales are significantly correlated at greater than $r = .30$. Furthermore, we see a particularly strong correlation between the 'fairness' item on the two scales (relative to the correlations between the other parallel items).

The final analyses involved a follow-up survey that we conducted about three months after the initial survey. Our hope was to use those participants ($n = 234$) who provided contact information (for potential participation in focus groups) to gauge the persistence of the

Table 2. Pearson (*r*) correlations for treatment participants between Support for Student-Perception Survey and Support for Teacher-Perception Survey responses.

		Student-Perception Surveys					
		1	2	3	4	5	**6**
Teacher-Perception Surveys	1) Fair	.53	.29	.32	.36	.38	**.45**
	2) Useful	.27	.37	.21	.16	.25	**.30**
	3) Objective	.34	.25	.32	.28	.25	**.34**
	4) Others	.36	.29	.27	.33	.23	**.35**
	5) Good idea	.36	.30	.30	.26	.34	**.38**
	6) Overall composite	**.53**	**.41**	**.41**	**.40**	**.41**	.52

Notes: 1) $N = 151–152$.
2) All correlations presented in the table above .20 are significant at the $p < .05$ level.

effects of the intervention. In this follow-up, we re-administered only the scale on student-perception surveys. About a third (32%) of the eligible participants responded. For those in the treatment group ($n = 31$), opinions remained consistent over this three-month span ($M_{pre} = 2.88$, SD = .91; $M_{post} = 2.90$, SD = .92). Their pre- and post-ratings also correlated with each other strongly ($r_{31} = .83$).

While a potentially encouraging sign for the endurance of our intervention, this result should be interpreted cautiously. The control teachers ($n = 44$) who completed both surveys became slightly more positive over the three-month span ($M_{pre} = 2.77$, SD = .67; $M_{post} = 2.86$, SD = .62) and showed less stability in their opinions between these pre- and post-assessments ($r_{44} = .46$).

Further analyses showed evidence of differential rates of volunteering for the follow-up survey. The subgroup of treatment participants who completed both surveys was similar to the original treatment group ($M_{\text{pre - original}} = 2.83$, SD = .85 versus $M_{\text{pre - both}} = 2.88$, SD = .91). However, the subgroup of control participants who completed both surveys was not representative of the original control group ($M_{\text{pre - original}} = 2.60$, SD = .77 versus $M_{\text{pre - both}} = 2.77$, SD = .67). In addition to the small sample sizes for these follow-up analyses, we found sampling differences between the participants who participated in both surveys (as compared to the composition of the original sample) and differences in consistency of opinions over time for the two groups – all of which make interpretation of these results challenging.

Discussion

Through a modest, dissonance-based intervention, we find that asking teachers about student-perception surveys in different ways can affect teachers' receptivity to this evaluative practice. Specifically, we find that juxtaposing questions on the viability of teachers evaluating administrators with questions about the viability of students evaluating teachers makes teachers more receptive to student-perception surveys as a component of their evaluation (as compared to directly asking them about the viability of student-perception surveys). Although the effect size of this intervention was modest, effect sizes should be calibrated with respect to the magnitude of the intervention (Cumming, 2014). In this case, the intervention was exceedingly brief (less than two minutes for most participants) and simple to execute.

Despite being more suggestive in nature, the exploratory analyses provide additional signals that participants' responses on these surveys comport with what one would expect.

NNSTOY teachers are more open to student-perception surveys than their colleagues who have not received the same recognition. Teachers of younger students view this evaluative practice with less enthusiasm than their colleagues who teach older students. These two findings accord with the logic that (a) teachers who have received positive reinforcement about their performance may be less apprehensive about being evaluated by students and that (b) teachers intuit that older students are more capable of providing fair, objective, potentially useful feedback. Because these explanations are speculative – our data do not speak directly to either finding – these results offer potential avenues for future study.

In Table 1a, we also saw signs that the intervention affected certain aspects of teachers' perceptions of student-perception surveys more than others. Specifically, the intervention did not appear to affect teachers' perceptions of the utility of student feedback or their concerns about students' objectivity. Instead, it appears that the intervention most affected teachers' perceptions of fairness and whether student-perception surveys were a good idea.

This finding also helps rule out an alternative explanation that a mere ordering effect caused the results. For instance, in 'anchoring' (Dillman et al., 2014), respondents answer subsequent items with similar ratings as an initial item because of the standard that is brought to mind by the initial item; in 'anchoring and adjusting' (Gehlbach & Barge, 2012) respondents answer similar adjacent items with similar ratings. However, neither of these potential explanations seem viable given that the intervention affected some items but not others.

Our next analyses sought to provide additional evidence regarding whether cognitive dissonance seemed plausible as the explanatory mechanism. Identifying a causal mechanism is inherently a speculative endeavour – for our research design, it is probably more reasonable to expect to learn about the effects of causes rather than the causes of effects (Bullock, Green, & Ha, 2010; Holland, 1986). In other words, for experimental designs such as ours it is easy to articulate how groups differ on particular outcomes; describing which part of the intervention is responsible for causing that difference cannot be done with the same precision. With this caveat in mind, our data are congruent with a cognitive dissonance explanation. More specifically, we find that treatment participants' responses on the two scales covary (at both the item and scale levels). Had we found no correlation between the responses on the scales, it would be hard to imagine that the cognitive dissonance from the juxtaposition of the scales caused the responses on the second scale to be higher. In addition, the correlations were particularly strong for the fairness item – an item likely to engender implicit comparisons between the student- and teacher-perception surveys.

Our attempts to ascertain whether the effects of the intervention endured over time were somewhat frustrated. Only a modest proportion of our original participants responded. These respondents may have been reasonably representative of the larger treatment group. Yet, it appeared that the control group of follow-up respondents were not representative of the original control group. Specifically, they held much more favourable initial views about student-perception surveys as compared to the overall control group. Furthermore, the control group showed much greater fluctuation in their opinions over these three months than their peers in the treatment group. All of these factors muddy our attempts to understand the persistence of the intervention. However, our attempt to gauge persistence was not devoid of information. Given the brief nature of the intervention, it would hardly have been surprising if the treatment effects had disappeared over time (Rogers & Frey, 2015). However, we find no evidence that the more positive attitudes of those in the treatment condition drifted back to baseline. Thus, while we are reticent to make a strong claim that the effects endured, we can produce no evidence that they faded either. Consequently, assessing the longevity of these effects seems like an especially important area for future research.

Finally, our study helps shed new light on a current debate in the cognitive dissonance literature: Does behaviour need to be implicated for dissonance to occur, or can dissonance result merely from incongruous cognitions that have no action implications (Brehm, 2007; Harmon-Jones et al., 2015)? Past studies on the 'even-handedness effect' (Dillman et al., 2014) suggest that, in at least some cases, dissonance can occur without implications for a respondent's behaviour. However, these studies asked respondents about topics that they were unlikely to have thought about much and that were largely irrelevant to their personal lives (i.e. freedoms for communist vs. western reporters in one example). We asked teachers about a topic of clear personal relevance, but which lacked clear action implications for them. Because of the clear personal relevance, one might have anticipated that their attitudes might be more deeply held, and thus more resistant to change simply by being brought into conflict with another cognition. Yet, our study finds that the treatment group still shifted their attitude towards student-perception surveys relative to the control group. This finding provides additional evidence congruent with the notion that cognitive dissonance may occur through conflicting cognitions alone; action implications may not always be necessary.

Limitations

In addition to the problems that arose in our attempts to learn about the duration of the effects of the intervention, other limitations of the study are important to weigh. One obvious issue is that the study provides only minimal evidence about what the mediating mechanism might be. Our theory is that participants in the treatment group have different attitudes towards student-perception surveys because they experienced a form of cognitive dissonance. However, other explanations may well be plausible and additional evidence to support (or disconfirm) our explanation would clearly strengthen our study.

Perhaps the most prominent question is the extent to which the sample might affect the validity of the findings. One version of this question revolves around internal validity. Does having a high proportion of nationally recognised teachers (and their friends and colleagues) in the sample jeopardise the integrity of the intervention? All participants were randomly assigned to condition, random assignment appeared to work (so far as we could check it), and we controlled for NNSTOY status. As a result of these checks and safeguards, we cannot come up with a plausible story as to how the internal validity might be threatened by the sample.

The second question is whether the sample affected the external validity or generalisability of the results. This possibility seems more concerning. Relative to a nationally representative sample of US teachers, our sample was more accomplished. While the level of accomplishment is clear for the NNSTOY teachers in our sample, it seems possible that the colleagues and associates of these teachers are also stronger and/or more experienced teachers than typical US teachers. As such, many teachers in our sample may have received more positive reinforcement about their teaching over the years than typical teachers. As a result, teachers in our sample might be more open to student-perception surveys as a component of how they are evaluated. So one potential threat to external validity is that a more typical population of teachers would be so averse to the use of student-perception surveys that a modest intervention such as this one could not possibly work.

On the other hand, an equally compelling story might be told that NNSTOY teachers (and their colleagues) are sufficiently confident in their teaching capacities, that they are relatively unafraid of student-perception surveys as an evaluation component. Consequently, the effects of the intervention may have been muted on this relatively elite sample of teachers. In this case, the threat to validity would be that the effects of our intervention would be stronger on a more typical population of teachers than the effects found in this study. The range of scores for each item and the overall Support for Student-Perception Surveys composite all extended from 1 to 5. This suggests that we did obtain a diverse sample of teachers with respect to their views on student-perception surveys. Presumably some of them are relatively representative of a more typical sample of US teachers. However, like almost all studies, the real test for the external validity of this study lies in replication attempts with varied samples.

Implications

With these limitations in mind, we want to be appropriately cautious about the potential implications of this study. However, assuming that the intervention could be replicated on future populations of teachers, we think these findings raise two especially intriguing possibilities. The first is a practical policy consideration. If school administrators wish to nudge their teachers to be more open regarding student-perception surveys, they may want to consider whether teachers should have opportunities to evaluate administrators. If future research suggests that the intervention worked, in part, because of a norm of evenhandedness (Dillman et al., 2014) or reciprocity (Cialdini, 2009), expanding the scope of these types of evaluations seems reasonable to entertain. A number of businesses have employed '360 degree evaluations' – a system in which any given individual receives feedback from subordinates, peers, and managers – as part of a cultural norm in their organisations (Peiperl, 2001). Perhaps schools might benefit from a similar approach.

Second, we think our findings signal some promise for the use of surveys as interventions. While typically thought of as data collection tools, surveys can be used to shift respondents' attitudes and beliefs. At times, surveys-as-interventions have been used with nefarious intentions, particularly in politics. The practice of push-polling consists of setting up a fraudulent poll in which a large number of respondents are typically asked a relatively small number of questions about a single candidate or issue where the questions are uniformly negative (AAPOR, 2007). The intent of these 'polls' is not to collect data but rather to push the opinions of voters by sowing seeds of doubt about particular candidates or issues. Other instances of surveys-as-interventions have been for more neutral purposes – e.g. to illustrate response order effects in survey design as described in the introduction. However, we think that surveys as interventions might be used to positively impact educational outcomes. The present study serves as a proof of concept for one such instance – our intervention shows how support might be generated for particular school policies. Providing individuals with feedback from surveys offers a related type of intervention that also may yield positive benefits for educational settings (Gehlbach et al., 2016). Thus, there may be future possibilities for scholars to use surveys as interventions that might help facilitate desired educational outcomes.

Conclusion

To our knowledge, this study is the first of its kind to leverage a survey as an intervention to shift teachers' beliefs – in this case, about the viability of using student-perception surveys as a component of their evaluation system. While much remains to be learned about the efficacy of this particular intervention – with respect to other populations of teachers and to the longevity of the effects – the basic approach offers some new ways to think about constructing interventions in education.

We expect that some school leaders might perceive a technique such as this to be too 'manipulative' for their tastes. Though they may be reluctant to use this survey approach in their own schools, perhaps they may still perceive potential benefits from employing 360-degree evaluations. Other school leaders will likely view this survey as no more manipulative than the array of positive and negative reinforcers already used in schools (e.g. linking teachers' pay with their students' standardised test scores as a means to bolster teachers' effort, or giving students extra recess for good behaviour). Perhaps school leaders might even use this intervention directly – for example, by having teachers complete a survey similar to the treatment group's version prior to a faculty meeting where the school's evaluation system is under discussion. They might use this approach to begin a conversation around the costs and benefits of implementing a more comprehensive evaluation system for all school personnel.

Thus, we presume that employing an intervention such as this one will be more appealing to some school leaders than others. However, we also hope that this type of survey-as-intervention approach sparks some creative new developments in how researchers think about improving an array of educational outcomes.

Acknowledgements

The authors are grateful to the Katherine Bassett, Bob Williams and their team at the National Network of State Teachers of the Year for their tremendous support in conducting this study and thoughtful comments on a draft of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Gevirtz Graduate School of Education.

ORCID

Hunter Gehlbach http://orcid.org/0000-0002-2852-2666

References

AAPOR. (2007). AAPOR Statements on "Push" Polls. Retrieved from https://www.aapor.org/Education-Resources/Resources/AAPOR-Statements-on-Push-Polls.aspx

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–125.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., … Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper #278). Retrieved from https://search.proquest.com/docview/860368237?accountid=14522

Berliner, D. C. (2013). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record, 115*(12), 1–26.

Brehm, J. W. (2007). A brief history of dissonance theory. *Social and Personality Psychology Compass, 1*, 381–391. doi:10.1111/j.1751-9004.2007.00035.x

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality & Social Psychology, 98*, 550–558.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (17699). Washington, DC: National Bureau of Economic Research.

Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Boston, MA: Pearson.

The Colorado Education Initiative. (2015). Using Student Perception Survey Results in Educator Evaluations. Retrieved from https://www.coloradoedinitiative.org/our-work/professional-learning/improving-success-for-all-students-toolkit/using-sps-results-in-educator-evaluations/

Cromidas, R. (2012). Survey of students about student surveys yields mixed opinions. Retrieved from https://ny.chalkbeat.org/2012/12/10/survey-of-students-about-student-surveys-yields-mixed-opinions/

Cumming, G. (2014). The new statistics. *Psychological Science, 25*, 7–29. doi:10.1177/0956797613504966

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.

Decker, G. (2012). Student surveys seen as unlikely evaluations element, for now. Retrieved from https://ny.chalkbeat.org/2012/11/28/student-surveys-seen-as-unlikely-addition-to-evaluations-for-now/

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: Wiley.

Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan, 94*, 24–28.

Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*, 93–106.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*, 195–202. doi:10.1037/h0023552

Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social Cognition, 30*, 652–668. doi:10.1521/soco.2012.30.6.652

Gehlbach, H. (2015). Seven survey sins. *The Journal of Early Adolescence, 35*, 883–897. doi:10.1177/0272431615578276

Gehlbach, H., & Barge, S. (2012). Anchoring and adjusting in questionnaire responses. *Basic and Applied Social Psychology, 34*, 417–433. doi:10.1080/01973533.2012.711691

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15*, 380–387. doi:10.1037/a0025704

Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher–student relationships and academic achievement. *Journal of Educational Psychology, 108*, 342–352. doi:10.1037/edu0000042

Gehlbach, H., & Robinson, C. (manuscript under review). Mitigating illusory results through pre-registration in education.

Good, T. L. (2014). What do we know about how teachers influence student performance on standardized tests: And why do we know so little about other student outcomes? *Teachers College Record, 116*(1), 1–41.

Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An action-based model of cognitive-dissonance processes. *Current Directions in Psychological Science, 24*, 184–189. doi:10.1177/0963721414566449

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–960.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: Bill & Melinda Gates Foundation.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*, 935–937. doi:10.1126/science.1238056

Kraft, M. A., & Gilmour, A. F. (2016). *Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness* (Working Paper). Brown University.

Martinie, M. A., Milland, L., & Olive, T. (2013). Some theoretical considerations on attitude, arousal and affect during cognitive dissonance. *Social and Personality Psychology Compass, 7*, 680–688. doi:10.1111/spc3.12051

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. doi:10.1037/0003-066X.50.9.741

MET Project. (2012). *Asking student about teaching*. Retrieved from https://www.metproject.org/downloads/Asking_Students_Practitioner_Brief.pdf

National Center for Educational Statistics. (2013). Digest of education statistics. Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_209.10.asp

Peiperl, M. A. (2001). Getting 360 degrees feedback right. *Harvard Business Review, 79*, 142–147, 177.

Rogers, T., & Frey, E. (2015). Changing behavior beyond the here and now. In *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 723–748).

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22*, 1359–1366. doi:10.1177/0956797611417632

Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy, 11*, 340–359.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*, 3628–3651. doi:10.1257/aer.102.7.3628

TEAMTN. (2015). Tennessee educator acceleration model: A Tennessee department of education website. Overview. Retrieved from https://team-tn.org/evaluation/overview/

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*, 26–30.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*, 1447–1451. doi:10.1126/science.1198364

Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal and self-ratings in 360" feedback® for teacher evaluation. *Journal of Personnel Evaluation in Education, 14*, 179–192. doi:10.1023/A:1008158904681

Appendix 1.

Descriptions of the measures used in this study:

The 5-item Support for Student-Perception Surveys scale:

How fair is it for student-perception surveys to be one of the sources of information in assessing your teaching performance?	Not fair at all	Mildly fair	Moderately fair	Quite fair	Extremely fair
How useful is it for you to receive feedback on your teaching from your students?	Not at all useful	Mildly useful	Moderately useful	Quite useful	Extremely useful
How objectively can your students assess your teaching performance?	Not at all objectively	Mildly objectively	Moderately objectively	Quite objectively	Extremely objectively
How supportive do you think *other teachers* are of using student-perception surveys to assess teaching performance?	Not at all supportive	Mildly supportive	Moderately supportive	Quite supportive	Extremely supportive
Overall, to what extent is it a good idea to have teachers' performance reviews be partially based on student input?	Not a good idea at all	A mildly good idea	A moderately good idea	Quite a good idea	An extremely good idea

The 5-item Support for Teacher-Perception Surveys scale:

How objectively can teachers evaluate their administrators?	Not at all objectively	Mildly objectively	Moderately objectively	Quite objectively	Extremely objectively
How fair is it for teacher-perception surveys to be one of the sources in assessing the performance of school administrators?	Not fair at all	Mildly fair	Moderately fair	Quite fair	Extremely fair
How supportive do you think *other teachers* are of using teacher-perception surveys to assess administrators' performance?	Not at all supportive	Mildly supportive	Moderately supportive	Quite supportive	Extremely supportive
How useful is it for administrators to receive feedback on their job performance from their faculty?	Not at all useful	Mildly useful	Moderately useful	Quite useful	Extremely useful
Overall, to what extent is it a good idea for administrators' evaluations to be based partially on teacher input?	Not a good idea at all	A mildly good idea	A moderately good idea	Quite a good idea	An extremely good idea

Note: For each item, the response options were scored on a 1-through-5 system where 1 = 'Not at all' and 5 = 'Extremely'